PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/40, 15/48, C07K 14/16

A1

(11) International Publication Number: WO 95/04147

(43) International Publication Date: 9 February 1995 (09.02.95)

(21) International Application Number: PCT/US94/05166

(22) International Filing Date: 10 May 1994 (10.05.94)

(30) Priority Data:
08/101,536 2 August 1993 (02.08.93) US

(71) Applicant: CHIRON CORPORATION [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US).

(72) Inventor: SELBY, Mark; 136 Galewood Circle, San Fransisco, CA 94131 (US).

(74) Agents: McCLUNG, Barbara, G. et al.; Chiron Corporation, Intellectual Property Dept. - R440, P.O. Box 8097, Emeryville, CA 94662-8097 (US).

(81) Designated States: AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, ES, FI, GB, GE, HU, JP, KG, KP, KR, KZ, LK, LU, LV, MD, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SI, SK, TJ, TT, UA, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: RECOMBINANT CONSTRUCTS USING REPLACEMENT SEQUENCES IN HYPERVARIABLE REGIONS

(57) Abstract

Novel recombinant constructs bearing immunogenic genes derived from RNA viruses are disclosed. The constructs contain gene sequences derived from immunogenic viral proteins (termed "backbone proteins" herein) which have hypervariable regions. The hypervariable regions are replaced, in whole or part, with consensus sequences or with corresponding hypervariable gene sequences from viral isolates other than the parent isolate. The constructs are conveniently engineered so that the replacement sequence is flanked with unique restriction sites. In this way, a variety of immunogenic sequences, e.g., from related strains of viruses, can be easily inserted into, and excised from, the construct. The proteins derived from the constructs can be used in vaccine compositions and diagnostics for the prevention and diagnosis, respectively, of the virus in question.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

BNSDOCID: <WO___9504147A1_I_>

RECOMBINANT CONSTRUCTS USING REPLACEMENT SEQUENCES IN HYPERVARIABLE REGIONS

Background of the Invention

Technical Field

The present invention relates generally to the recombinant production of proteins. More particularly, the invention relates to recombinant constructs derived from RNA viruses having hypervariable regions, for use in the production of vaccines and diagnostics.

Background of the Invention

RNA viruses have highly variable genomes, likely due to the increased incidence of spontaneous mutations caused by the poor fidelity of RNA replicases and reverse transcriptase. As a result, these viruses exhibit a large amount of sequence diversity from isolate to isolate, clustered in regions known as hypervariable domains. Consequently, antibodies raised against one viral isolate are frequently unable to effectively cross-neutralize related viral isolates.

For example the human immunodeficiency viruses, HIV-1 and HIV-2, display extensive variation in the envelope gene. The envelope gene of HIV-1 encodes a 160 kDa precursor glycoprotein, termed gp160, which is enzymatically cleaved into two glcoproteins, the extracellular protein gp120 and the transmembrane protein, gp41. The gp120 protein includes several hypervariable regions (V1-V5). The V3 region of gp120 from numerous isolates, including isolates belonging to all five of the currently identified subtypes of HIV-1, has been well characterized. Myers *et al.* Los Alamos Database, Los Alamos National Laboratory, Los Alamos, New Mexico (1992). This region consists of approximately 35 amino acids and is bounded by cysteines. The V3 region serves as the primary neutralization determinant of HIV-1 and as a major determinant of HIV-1 cell tropism. Hwang *et al. Science* (1991) 253:71-74. Chimeric HTLV-IIIB viruses, including portions of the HIV-1 V3 region, have been constructed in an attempt to further characterize the function of the region. Hwang *et al. Science*

(1991) 253:71-74; Hwang et al. Science (1992) 257:535-537. Compositions have been prepared using synthetic peptides from the V3 region. *See, e.g.*, PCT Publication Nos. WO93/03766 (published 4 March 1993), WO92/22572 (published 23 December 1992); EPA Publication No. 328,403 (published 16 August 1989). However, synthetic peptides often lack the activity of the native molecules.

Similarly, influenza A, B and C viruses are enveloped RNA viruses which have surface glycoproteins exhibiting much variability. Influenza A viruses have been isolated from a wide variety of animal species, including humans, and the surface glycoproteins therefrom exhibit a greater degree of variability than their B and C counterparts. The viral genome encodes two envelope proteins termed HA (hemagglutinin) and NA (neuraminidase). HA is essential for viral attachment and entry and mediates the initial attachment of the virus particle to its cellular receptor, a glycoconjugate terminating in a sialic acid residue. NA serves an enzymatic function and removes the terminal sialic acids from oligosaccharides on cell surface proteins and glycolipids. In an attempt to enhance the immunogenic potential of the virus, chimeric hemagglutinins have been constructed wherein a six amino acid loop region of a subtype 1 HA (termed "H1") was replaced with the corresponding region from subtype 2 HA (termed "H2") and subtype 3 HA (termed "H3"), respectively. The H1-H3 chimera was reactive with both of the subtypes of the virus. Li *et al. J. Virol.* (1992) 66:309-404.

As with the above-described viruses, hepatitis C virus (HCV), the major causative agent of post-transfusion Non-A, Non-B hepatitis (NANBH), is an RNA virus showing sequence diversity from isolate to isolate. Several hypervariable domains have been described and, like HIV, the putative viral envelope proteins, termed E1 and E2/NS1, respectively, show substantial amino acid sequence variation between the groups. Specifically, a hypervariable region, located at the amino terminus of E2/NS1, has been identified. Weiner *et al. Virology* (1991) 180:842-848. This region occurs between amino acids 384-414, using the amino acid numbering system of HCV-1. International Publication No. WO93/06126 (published 1 April 1993) describes the use of HCV epitopes found within this region in vaccine compositions. Kato *et al. J.*

Virol. (1993) 67:3923-3930 describes the production of a humoral immune response to the E2/NS1 region (termed HVR1 therein) derived from a Japanese HCV isolate.

However, none of the above-described art provides recombinant constructs that facilitate the replacement of an existing hypervariable region from a particular isolate with a sequence which will cross-react with a number of isolates. Accordingly, there remains a need for proteins that maintain the activity of the native molecules and which are immunologically cross-reactive with multiple isolates of the particular RNA virus in question.

Disclosure of the Invention

The present invention is based on the production of viral constructs containing a wild-type backbone protein in which at least a portion of one hypervariable region therein has been substituted with either a corresponding consensus sequence or a sequence from a corresponding hypervariable region found in a different isolate. The replacement sequences are bounded by unique restriction sites in order to facilitate insertion and excision of the replacement sequences.

Accordingly, in one embodiment, the invention is directed to a recombinant construct comprising a nucleotide sequence encoding a backbone immunogenic protein derived from a parental RNA virus. The backbone protein is characterized as having at least one wild-type hypervariable region in its native state. A sequence from the hypervariable region is substituted with a replacement sequence corresponding thereto and the replacement sequence is also flanked at the 3'- and 5'-ends thereof with unique restriction sites.

In another embodiment, the invention is directed to a recombinant vector comprising:

(a) a DNA sequence encoding a human immunodeficiency virus type 1 (HIV-1) subtype B gp120 wherein the wild-type V3 hypervariable region is replaced with a DNA sequence encoding a corresponding consensus sequence for HIV-1 subtype B V3; and

(b) control sequences that are operably linked to the DNA sequence whereby the DNA sequence can be transcribed and translated in a host cell and at least one of the control sequences is heterologous to the DNA sequence.

In still another embodiment, the invention is directed to a recombinant vector comprising:

(a) a DNA sequence encoding a human immunodeficiency virus type 1 (HIV-1) subtype E gp120 wherein the wild-type V3 hypervariable region is replaced with a DNA sequence encoding a corresponding consensus sequence for HIV-1 subtype E V3; and

(b) control sequences that are operably linked to the DNA sequence whereby the sequence can be transcribed and translated in a host cell and at least one of the control sequences is heterologous to the DNA sequence.

In other embodiments, the invention is directed to plasmids $pCMVgp120_{SF2.3/NAE.C}$ (ATCC No. 69365), $pCMVgp120_{SF2.3/NT9142}$ (ATCC No. 69366), and $pCMVgp120_{CM235.3/NT9142}$ (ATCC No. 69367).

In still other embodiments, the invention is directed to host cells transformed with these constructs and methods of producing recombinant polypeptides using the transformed cells.

In another embodiment, the invention is directed to a human immunodeficiency virus (HIV) gp120 wherein the wild-type V3 sequence is replaced in whole or part with a consensus sequence corresponding thereto.

In yet other embodiments, the invention is directed to vaccine compositions and diagnostics comprising the engineered gp120 sequences, as well as methods of using the same.

These and other embodiments of the subject invention will readily occur to those of skill in the art in light of the disclosure herein.

Brief Description of the Figures

Figure 1 shows the wild-type nucleotide sequence and corresponding amino acid sequence for SF2 gp120. The V3 region is underlined, and the mature gp120 is also indicated (wherein the signal sequence is lacking).

Figures 2A-E depict the V3 consensus sequences for the five HIV-I subtypes. Figure 2A shows the consensus sequences for subtype A. Figure 2B depicts the consensus sequences for subtype B; Figure 2C shows the consensus sequences for subtype C; Figure 2D depicts the consensus sequences for subtype D; and Figure 2E depicts the consensus sequences for subtype E. (Myers *et al.* Los Alamos Database, Los Alamos National Laboratory, Los Alamos, New Mexico (1992)).

Figure 3 shows the subtype B V3 consensus sequence and flanking 5' and 3' sequences present in plasmid pCMVgp120$_{SF2.3/NAE.C}$ which include the nucleotide changes to provide for the unique *Nru*I and *Xba*I sites. The V3 consensus sequence is underlined. Δ denotes nucleotide changes made to generate the unique *Nru*I site; □ denotes the nucleotide changes made to generate the unique *Xba*I site; and ○ indicates the nucleotide changes made to conform the SF2 V3 nucleotide sequence to the subtype B consensus sequence.

Figure 4 is a diagram of the V3 loop region, showing the flanking NruI and XbaI sites and the BglII and Bsu361 sites, further upstream and downstream, respectively. The elevated bars denote the nucleotides mutated in order to generate the V3 subtype B consensus sequence. The horizontal bars denote the overlapping oligonucleotides used togenerate the consensus sequence.

Figure 5 is a diagram of plasmid pCMV6a120-SF2.

Figure 6 shows the HCV E2/NS1 hypervariable regions for 90 HCV isolates and the determined consensus sequence based on these sequences.

Detailed Description of the Invention

The practice of the present invention will employ, unless otherwise indicated, conventional methods of virology, microbiology, molecular biology and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature. See, *e.g.*, Sambrook, *et al. Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); Maniatis *et al. Molecular Cloning: A Laboratory Manual* (1982); *DNA Cloning: A Practical Approach*, vol. I & II (D. Glover, ed.); *Oligonucleotide Synthesis* (N. Gait, ed., 1984); *Nucleic Acid Hybridization* (B. Hames & S. Higgins, eds., 1985); *Transcription and Translation* (B. Hames & S. Higgins,

eds., 1984); *Animal Cell Culture* (R. Freshney, ed., 1986); Perbal, *A Practical Guide to Molecular Cloning* (1984).

All publications, patents and patent applications cited herein, whether *supra* or *infra*, are hereby incorporated by reference in their entirety.

I. Definitions

In describing the present invention, the following terms will be employed, and are intended to be defined as indicated below.

By "RNA virus" is meant any virus having an RNA genome. An RNA virus can be made up of single or double stranded RNA. Such viruses include, without limitation, members of the families Picornaviridae (*e.g.*, polioviruses, hepatitis A virus, etc.); Caliciviridae; Togaviridae (*e.g.*, rubella virus, etc.); Flaviviridae (*e.g.*, hepatitis C virus); Coronaviridae; Reoviridae; Birnaviridae; Rhabodoviridae (*e.g.*, rabies virus, etc.); Filoviridae; Paramyxoviridae (*e.g.*, mumps virus, measles virus, respiratory syncytial virus, etc.); Orthomyxoviridae (*e.g.*, influenza virus types A, B and C, etc.); Bunyaviridae; Arenaviridae; Retroviradae (*e.g.*, HTLV-I; HTLV-II; HIV-1 (also known as HTLV-III, LAV, ARV, hTLR, etc.), including but not limited to the isolates HIV_{IIIb}, HIV_{SF2}, HIV_{LAV}, HIV_{LAI}, HIV_{MN}); HIV-2; simian immundeficiency virus (SIV); as well as RNA viruses not yet classified into the above families, such as hepatitis delta virus and hepatitis E virus, among others. See, *e.g. Virology*, 3rd Edition (W.K. Joklik ed. 1988); *Fundamental Virology*, 2nd Edition (B.N. Fields and D.M. Knipe, eds. 1991), for a description of RNA viruses. Additionally, the present invention may have application to DNA viruses as well.

By "hypervariable" region or domain is meant a region showing a consistent pattern of amino acid variation between at least two particular viral isolates or subpopulations. A hypervariable domain can vary from isolate to isolate by as little as one amino acid and will include at least one epitope. Such hypervariable domains can be identified by comparing amino acid sequences between viral isolates by *e.g.*, aligning the conserved domains of two or more isolates using computer programs, such as ALIGN 1.0, available from the University of Virginia, Department of Biochemistry (Attn: Dr. William R. Pearson). *See*, Pearson *et al. Proc. Natl. Acad. Sci. USA*

85:2444-2448. It is to be understood that the amino acid numbers given for a particular hypervariable region are somewhat subjective and a matter of choice. Thus, the beginning and end of a particular hypervariable domain are approximate and include overlapping domains or subdomains. This definition encompasses domains designated as "variable."

By "backbone polypeptide" is meant the wild-type amino acid sequence of a particular immunogenic viral protein which includes one or more hypervariable domains as described above. It is to be understood that the backbone proteins of the present invention may have, and generally do include, more than one hypervariable region. Accordingly, when the hypervariable region is referred to in the singular, the term is meant to encompass one or more of such regions. A "backbone gene" is a gene encoding a backbone polypeptide.

By "wild-type sequence" is meant an amino acid or nucleotide sequence, respectively, which corresponds to the primary sequence recovered from a viral isolate occurring in nature. Thus, the term "wild-type gp120" denotes a gp120 having a primary sequence equivalent to the sequence of naturally occurring gp120 found in the HIV isolate in question. The wild-type sequence need not be physically isolated from the virus but may be produced synthetically or recombinantly.

By "native polypeptide" is meant a polypeptide having a conformation equivalent to the conformation of the polypeptide as it occurs in nature.

By "replacement sequence" is meant an amino acid or nucleotide sequence, respectively, which is used in place of the wild-type sequence occurring at a particular region in the backbone sequence.

The terms "polypeptide" and "protein" refer to a polymer of amino acid residues and are not limited to a minimum length of the product. Thus, peptides, oligopeptides, dimers, multimers, and the like, are included within the definition. Both full-length proteins and fragments thereof are encompassed by the definition. The terms also include postexpression modifications of the polypeptide, for example, glycosylation, acetylation, phosphorylation and the like.

A polypeptide is "immunologically reactive" when it includes one or more epitopes and elicits antibodies that react with the native protein. These antibodies

may also neutralize infectivity, and/or mediate antibody-complement or antibody dependent cell cytotoxicity to provide protection to an immunized host. Immunological reactivity may be determined in a standard immunoassay, such as a competition assay, as is known in the art.

By "unique restriction site" is meant a site of a vector or nucleotide sequence which is cleaved by a particular restriction endonuclease which does not substantially cleave another site in the vector or nucleotide sequence. Two such sites will be present flanking the engineered gene containing the hypervariable region replacement sequence of the present invention. It is to be understood that the unique restriction sites flanking the gene of interest can be the same -- *i.e.*, the same restriction enzyme can act at both of the flanking sequences, so long as it does not cleave at other sites within the vector or nucleotide sequence.

"Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. A regulatory element "operably linked" to a structural sequence is ligated in such a way that expression of the structural sequence is achieved under conditions compatible with the regulatory elements.

"Recombinant" as used herein to describe a polynucleotide means a polynucleotide of genomic, cDNA, semisynthetic, or synthetic origin which, by virtue of its origin or manipulation: (1) is not associated with all or a portion of the polynucleotide with which it is associated in nature; and/or (2) is linked to a polynucleotide other than that to which it is linked in nature. The term "recombinant" as used with respect to a protein or polypeptide means a polypeptide produced by expression of a recombinant polynucleotide. "Recombinant host cells," "host cells," "cells," "cell lines," "cell cultures," and other such terms denoting prokaryotic microorganisms or eucaryotic cell lines cultured as unicellular entities, are used inter-changeably, and refer to cells which can be, or have been, used as recipients for recombinant vectors or other transfer DNA, and include the progeny of the original cell which has been transfected. It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement to the original parent, due to accidental or deliberate mutation. Progeny of

the parental cell which are sufficiently similar to the parent to be characterized by the relevant property, such as the presence of a nucleotide sequence encoding a desired peptide, are included in the progeny intended by this definition, and are covered by the above terms.

A "control element" refers to a polynucleotide sequence which effects the expression of a coding sequence to which it is linked. The term includes promoters, terminators, and when appropriate, leader sequences and enhancers.

"Transformation," as used herein, refers to the insertion of an exogenous polynucleotide into a host cell, irrespective of the method used for insertion: for example, direct uptake, transduction, or f-mating. The exogenous polynucleotide may be maintained as a nonintegrated vector, for example, a plasmid, or alternatively, may be integrated into the host genome.

A "vector" is a replicon in which a heterologous polynucleotide segment is attached, so as to bring about the replication and/or expression of the attached segment, such as a plasmid, transposon, phage, etc.

As used herein, "treatment" refers to any of (i) the prevention of infection or reinfection, as in a traditional vaccine, (ii) the reduction or elimination of symptoms, and (iii) the substantial or complete elimination of the virus. Treatment may be effected prophylactically (prior to infection) or therapeutically (following infection).

As used herein, a "biological sample" refers to a sample of tissue or fluid isolated from an individual, including but not limited to, for example, blood, plasma, serum, fecal matter, urine, bone marrow, bile, spinal fluid, lymph fluid, samples of the skin, external secretions of the skin, respiratory, intestinal, and genitourinary tracts, tears, saliva, milk, blood cells, tumors, organs, biopsies and also samples of *in vitro* cell culture constituents including but not limited to conditioned media resulting from the growth of cells and tissues in culture medium, *e.g.*, MAb producing myeloma cells, recombinant cells, and cell components.

By "individual" is meant any member of the subphylum chordata, including, without limitation, humans and other primates, including non-human primates such as chimpanzees and other apes and monkey species; farm animals such

as cattle, sheep, pigs, goats and horses; domestic mammals such as dogs and cats; laboratory animals including rodents such as mice, rats and guinea pigs; birds, including domestic, wild and game birds such as chickens, turkeys and other gallinaceous birds, ducks, geese, and the like. The term does not denote a particular age. Thus, both adult and newborn individuals are intended to be covered.

II. Modes of Carrying Out the Invention

The present invention is based on the development of novel recombinant constructs that encode immunogenic polypeptidess derived from RNA viruses. In particular, recombinant constructs and expression vectors are disclosed which contain gene sequences derived from immunogenic viral proteins (termed "backbone proteins" herein) which have hypervariable regions. The hypervariable regions in the constructs of the present invention are replaced, in whole or part, with consensus sequences or with corresponding hypervariable gene sequences from viral isolates other than the parent isolate. The constructs are conveniently engineered so that the replacement sequence, when present in the vector, is flanked with unique restriction sites. In this way, a variety of immunogenic sequences, *e.g.*, from related strains of viruses, can be easily inserted into, and excised from, the construct. Thus, the technique is particularly useful with viruses having high mutation rates and which therefore exist in highly variant forms.

The present system has been exemplified herein with respect to hypervariable regions present in the gp120 envelope protein of HIV-1 and particularly with constructs derived from HIV-1 subtypes B and E. However, it is to be understood that the invention is equally applicable to other immunogenic proteins derived from HIV, such as the envelope proteins gp41 and gp160, as well as gag and pol, and to a wide variety of other RNA viruses, such as those discussed above including, without limitation, HCV and the influenza viruses.

Proteins expressed with the vectors of the present invention can be used in vaccine compositions to provide protection against a wide variety of viral isolates. Similarly, the proteins can be used to produce antibodies and the proteins and/or

antibodies used as diagnostics for detecting the presence or absence of virus, or antibodies to the virus, in a biological sample.

As explained above, the invention utilizes one or more substituted hypervariable domains present in a larger, backbone polypeptide. The backbone polypeptide is one derived from an immunogenic viral protein and can include the full-length, truncated or mutated sequence of the wild-type polypeptide. However, the use of the full-length or near full-length wild-type sequence is preferred as the protein derived therefrom will tend to retain the native conformation and hence function as the native protein would.

The particular gene encoding a backbone polypeptide to be used is one which is known to be important in the immune response to the virus in question. The gene can be obtained from the viral isolate of interest using recombinant methods, such as by screening reverse transcripts of mRNA, by screening genomic libraries from cells expressing the gene, or by deriving the gene from a vector known to include the same. The gene can then be isolated for further use or, if already present in a suitable expression vector, be manipulated *in situ*, as described further below. The gene of interest can also be produced synthetically, rather than cloned. The nucleotide sequence can be designed with the appropriate codons for the particular amino acid sequence desired. In general, one will select preferred codons for the intended host in which the sequence will be expressed. The complete sequence is assembled from overlapping oligonucleotides prepared by standard methods and assembled into a complete coding sequence. *See, e.g.*, Edge (1981) *Nature* 292:756; Nambair *et al.* (1984) *Science* 223:1299; Jay *et al.* (1984) *J. Biol. Chem.* 259:6311.

For purposes of the present invention, the nucleotide sequence encoding either the entire hypervariable region (and, optionally, N- and C-terminal flanking regions), or a portion thereof, is replaced with a different, corresponding nucleotide sequence, such that the resulting transcript includes an immunogenic viral polypeptide.

The replacement sequence can comprise a consensus sequence, determined by comparing the hypervariable sequences of three or more viral isolates within a specific group and constructed using nucleotide sequences encoding the most

frequently encountered amino acids at a particular position. In this way, proteins can be engineered that cross-react with several different viral isolates in the group and thus can be used in broad spectrum vaccines, effective against several viral isolates. This technique is particularly useful with viruses such as HIV, for which a multitude of sequences have been determined and for which broad spectrum protection has been problematic.

Depending on the particular isolate targeted, the consensus sequence can vary from the wild-type parental sequence by as little as one base pair, effecting a change in a single amino acid, or can encompass several hundred base pair changes. These mutations can be made to the wild-type sequence using conventional techniques such as by preparing synthetic oligonucleotides including the mutations and inserting the mutated sequence into the gene encoding a backbone polypeptide using restriction endonuclease digestion. (See, *e.g.*, Kunkel, T.A. *Proc. Natl. Acad. Sci. USA* (1985) 82:448; Geisselsoder *et al. BioTechniques* (1987) 5:786.) Alternatively, the mutations can be effected using a mismatched primer (generally 10-20 nucleotides in length) which hybridizes to the wild-type nucleotide sequence (generally cDNA corresponding to the RNA sequence), at a temperature below the melting temperature of the mismatched duplex. The primer can be made specific by keeping primer length and base composition within relatively narrow limits and by keeping the mutant base centrally located. Zoller and Smith, *Methods Enzymol.* (1983) 100:468. Primer extension is effected using DNA polymerase, the product cloned and clones containing the mutated DNA, derived by segregation of the primer extended strand, selected. Selection can be accomplished using the mutant primer as a hybridization probe. The technique is also applicable for generating multiple point mutations. *See, e.g.*, Dalbie-McFarland *et al. Proc. Natl. Acad. Sci USA* (1982) 79:6409. PCR mutagenesis will also find use for effecting the desired mutations.

The hypervariable region can also be replaced, in whole or part, with corresponding sequences of any of various related isolates, rather than with consensus sequences. This is particularly useful for viruses which mutate rapidly, such as influenza viruses, for which new vaccine compositions must be reformulated on a regular basis. Thus, replacement sequences can be obtained (*e.g.*, either by direct

isolation, recombinantly or synthetically) from an isolate of interest and inserted into the backbone polypeptide in place of the already existing hypervariable domain, using techniques described above. The replacement sequence obtained in this manner will generally encode at least one epitope and will therefore usually comprise a sequence encoding a minimum of about five amino acids, more typically a minimum of about eight amino acids, and even more typically, a minimum of about 10 amino acids. If more than one epitope is present, the replacement region will be at least as big as the combined sequences of the epitopes. However, since epitopes can overlap, the minimum amino acid sequence for the replacement sequence may be less than the sum of the individual epitopes.

As explained above, the engineered sequence will be flanked with unique restriction sites such that it can easily be excised from an expression vector and other replacement sequences inserted. This allows for the substitution of one consensus sequence for another (i.e, one viral subtype consensus sequence can be substituted for another viral subtype consensus sequence in the backbone) or for the substitution of the existing hypervariable region in the backbone (either the wild-type or substituted region) with another hypervariable region from a related isolate. Thus, for example, an HIV-1 subtype B consensus sequence in a construct can be readily replaced with, for example a subtype E consensus sequence, so that vaccines or diagnostics can be easily formulated, depending on the group of viral isolates targeted. Thus, if for example, a vaccine composition or diagnostic was desired for use against the subtype B isolates, the subtype B consensus sequence would be inserted into the gp120 backbone, making use of the unique restriction sites, to provide for broad protection against most or all of the viral isolates in that subtype.

The design of such unique restriction sites is accomplished by either analyzing the sequence of the vector to be used for existing restriction sites, or by subjecting the expression vector and engineered sequence to a battery of different restriction endonucleases and determining which of the enzymes do not cleave the vector and nucleotide sequence. Sequences capable of being cleaved by the selected enzyme can then be added to the engineered gene for insertion into the vector. Alternatively, sequences already present in the vector that flank the engineered genes

can be mutated, using the techniques described above, to produce the unique restriction sites.

The gene sequence encoding the backbone polypeptide with the replacement hypervariable sequence, is inserted into an expression vector, using methods known to those of skill in the art. Alternatively, if the engineered gene sequence is already present in a suitable expression vector (*i.e.*, when mutagenesis has been done *in situ* to a gene encoding a backbone polypeptide already present in an existing vector), the vector can be used as is, without further manipulation.

More particularly, in order to obtain expression of the engineered sequence, host cells are transformed with expression vectors which include control sequences operably linked to the desired coding sequence. Suitable expression systems for use with the present invention include systems which function in eucaryotic and procaryotic host cells. Particularly useful eucaryotic hosts include mammalian, yeast and insect cells. Typical procaryotic hosts are bacterial and usually involve the use of *E. coli*.

The control sequences will be compatible with the particular host cell used. For example, typical promoters for mammalian cell expression include the SV40 early promoter, the CMV promoter, the mouse mammary tumor virus LTR promoter, the adenovirus major late promoter (Ad MLP), and the herpes simplex virus promoter, among others. Other non-viral promoters, such as a promoter derived from the murine metallothionein gene, will also find use in mammalian constructs. Mammalian expression may be either constitutive or regulated (inducible), depending on the promoter. Typically, transcription termination and polyadenylation sequences will also be present, located 3' to the translation stop codon. Examples of transcription terminator/polyadenylation signals include those derived from SV40 (Sambrook *et al.* (1989) "Expression of cloned genes in cultured mammalian cells" in *Molecular Cloning: A Laboratory Manual*). Introns, containing splice donor and acceptor sites, may also be designed into the constructs of the present invention.

Enhancer elements can also be used in the mammalian constructs to increase expression levels. Examples include the SV40 early gene enhancer (Dijkema et al. *EMBO J.* (1985) 4:761) and the enhancer/promoters derived from the long

terminal repeat (LTR) of the Rous Sarcoma Virus (Gorman et al. *Proc. Natl. Acad. Sci. USA* (1982b) 79:6777) and human cytomegalovirus (Boshart et al. *Cell* (1985) 41:521). A leader sequence can also be present which includes a sequence encoding a signal peptide, to provide for the secretion of the foreign protein in mammalian cells. Preferably, there are processing sites encoded between the leader fragment and the gene of interest such that the leader sequence can be cleaved either *in vivo* or *in vitro*. The adenovirus tripartite leader is an example of a leader sequence that provides for secretion of a foreign protein in mammalian cells.

Once complete, the mammalian expression vectors can be used to transform any of several mammalian cells. Methods for introduction of heterologous polynucleotides into mammalian cells are known in the art and include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei.

Mammalian cell lines available as hosts for expression are also known and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (e.g., Hep G2), as well as others.

The constructs of the present invention can also be expressed in yeast. Control sequences for yeast vectors are known and include promoters such as alcohol dehydrogenase (ADH) (EPO Publication No. 284,044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH), hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, and pyruvate kinase (PyK) (EPO Publication No. 329,203). The yeast *PHO5* gene, encoding acid phosphatase, also provides useful promoter sequences (Myanohara et al. *Proc. Natl. Acad. Sci. USA* (1983) 80:1). In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For example, upstream activating sequences (UAS) of one yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription

activation region (U.S. Patent Nos. 4,876,197 and 4,880,734). Other examples of hybrid promoters include promoters which consist of the regulatory sequences of either the *ADH2*, *GAL4*, *GAL10*, or *PHO5* genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as *GAP* or *PyK* (EPO Publication No. 164,556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription.

Other control elements which may be included in the yeast expression vectors are terminators (*e.g.*, from *GAPDH* and from the enolase gene (Holland *J. Biol. Chem.* (1981) 256:1385), and leader sequences which encode signal sequences for secretion. DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the yeast invertase gene (EPO Publication No. 012,873; JPO Publication No. 62,096,086) and the α-factor gene (U.S. Patent Nos. 4,588,684, 4,546,083 and 4,870,008; EPO Publication No. 324,274; PCT Publication No. WO 89/02463). Alternatively, leaders of non-yeast origin, such as an interferon leader, also provide for secretion in yeast (EPO Publication No. 060,057).

Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors have been developed for, *inter alia*, the following yeasts: *Saccharomyces cerevisiae* (Hinnen *et al. Proc. Natl. Acad. Sci. USA* (1978) 75:1929; Ito *et al. J. Bacteriol.* (1983) 153:163); *Saccharomyces carlsbergeneis*; *Candida albicans* (Kurtz *et al. Mol. Cell. Biol.* (1986) 6:142); *Candida maltosa* (Kunze *et al. J. Basic Microbiol.* (1985) 25:141); *Hansenula polymorpha* (Gleeson *et al. J. Gen. Microbiol.* (1986) 132:3459; Roggenkamp *et al. Mol. Gen. Genet.* (1986) 202:302); *Kluyveromyces fragilis* (Das *et al. J. Bacteriol.* (1984) 158:1165); *Kluyveromyces lactis* (De Louvencourt *et al. J. Bacteriol.* (1983) 154:737; Van den Berg *et al. Bio/Technology* (1990) 8:135); *Pichia guillerimondii* (Kunze *et al. J. Basic Microbiol.* (1985) 25:141); *Pichia pastoris* (Cregg *et al. Mol. Cell. Biol.* (1985) 5:3376; U.S. Patent Nos. 4,837,148 and 4,929,555); *Schizosaccharomyces pombe* (Beach and Nurse, *Nature* (1981) 300:706); and *Yarrowia lipolytica* (Davidow *et al. Curr. Genet.* (1985) 10:380471; Gaillardin *et al. Curr. Genet.* (1985) 10:49).

Methods of introducing exogenous DNA into yeast hosts are well known in the art, and typically include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations.

Bacterial expression systems can also be used with the present constructs. Control elements for use in bacteria include promoters, optionally containing operator sequences, and ribosome binding sites. Useful promoters include sequences derived from sugar metabolizing enzymes, such as galactose, lactose (*lac*) and maltose. Additional examples include promoter sequences derived from biosynthetic enzymes such as tryptophan (*trp*), the b-lactamase (*bla*) promoter system, bacteriophage lambda PL, and T5. In addition, synthetic promoters, such as the tac promoter (U.S. Patent No. 4,551,433), which do not occur in nature also function in bacterial host cells.

The foregoing systems are particularly compatible with *E. coli*. However, numerous other systems for use in bacterial hosts such as *Bacillus spp.*, *Streptococcus spp.*, and *Streptomyces spp.*, among others, are also known. Methods for introducing exogenous DNA into these hosts typically include the use of $CaCl_2$ or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation.

Other systems for expression of the engineered sequences include insect cells and vectors suitable for use in these cells. The systems most commonly used are derived from the baculovirus *Autographa californica* polyhedrosis virus (AcNPV). Generally, the components of the expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a convenient restriction site for insertion of the heterologous gene or genes to be expressed; a wild type baculovirus with a sequence homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the heterologous gene into the baculovirus genome); and appropriate insect host cells and growth media.

Promoters for use in the vectors are typically derived from structural genes, abundantly transcribed at late times in a viral infection cycle. Examples include sequences derived from the gene encoding the viral polyhedron protein, Friesen *et al.*

(1986) "The Regulation of Baculovirus Gene Expression" in: *The Molecular Biology of Baculoviruses* (ed. Walter Doerfler); EPO Publication Nos. 127,839 and 155,476; and the gene encoding the p10 protein Vlak *et al. J. Gen. Virol.* (1988) 69:765.

The plasmid usually also contains the polyhedrin polyadenylation signal (Miller *et al. Ann. Rev. Microbiol.* (1988) 42:177) and a procaryotic ampicillin-resistance (*amp*) gene and origin of replication for selection and propagation in *E. coli*. DNA encoding suitable signal sequences can also be included and is generally derived from genes for secreted insect or baculovirus proteins, such as the baculovirus polyhedrin gene (Carbonell *et al. Gene* (1988) 73:409), as well as mammalian signal sequences such as those derived from genes encoding human α-interferon, Maeda *et al. Nature* (1985) 315:592; human gastrin-releasing peptide, Lebacq-Verheyden *et al. Molec. Cell. Biol.* (1988) 8:3129; human IL-2, Smith *et al. Proc. Natl. Acad. Sci. USA* (1985) 82:8404; mouse IL-3, (Miyajima *et al. Gene* (1987) 58:273; and human glucocerebrosidase, Martin *et al. DNA* (1988) 7:99.

Currently, the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed. These include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32 basepairs downstream from the ATT; see Luckow and Summers, *Virology* (1989) 17:31). The desired DNA sequence is inserted into the transfer vector, using known techniques (*see*, Summers and Smith, *supra*; Ju *et al.* (1987); Smith *et al. Mol. Cell. Biol.* (1983) 3:2156; and Luckow and Summers (1989) and an insect cell host is cotransformed with the heterologous DNA of the transfer vector and the genomic DNA of wild type baculovirus--usually by cotransfection. The vector and viral genome are allowed to recombine. The packaged recombinant virus is expressed and recombinant plaques are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA ("MaxBac" kit). These techniques are generally known to those skilled in the art and fully described in Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987) (hereinafter "Summers and Smith").

Recombinant baculovirus expression vectors have been developed for infection into several insect cells. For example, recombinant baculoviruses have been developed for, *inter alia*: *Aedes aegypti, Autographa californica, Bombyx mori, Drosophila melanogaster, Spodoptera frugiperda*, and *Trichoplusia ni*.

It is often desirable that the polypeptides prepared using the above systems be fusion polypeptides. As with non-fusion proteins, these proteins may be expressed intracellularly or may be secreted from the cell into the growth medium.

Once expressed, the polypeptide including the replacement sequence can be isolated from the above-described host cells using any of several techniques known in the art. If the expression system secretes the protein into growth media, the protein can be purified directly from the media. If the protein is not secreted, it is isolated from cell lysates. The protein can then be further purified using techniques known in the art, such as column chromatography, HPLC, immunoadsorbent techniques, affinity chromatography and immunoprecipitation. Activity of the purified proteins can be determined using standard assays, based on specific properties of the various native proteins.

The proteins of the present invention or immunoreactive fragments thereof can be used to produce antibodies, both polyclonal and monoclonal. If polyclonal antibodies are desired, a selected mammal, (e.g., mouse, rabbit, goat, horse, etc.) is immunized with an antigen of the present invention, or its fragment, or a mutated antigen. Serum from the immunized animal is collected and treated according to known procedures. If serum containing polyclonal antibodies is used, the polyclonal antibodies can be purified by immunoaffinity chromatography, using known procedures.

If monoclonal antibodies are desired, these can be prepared using somatic cell hybridization techniques described initially by Kohler and Milstein, *Nature* (1975) 256:495-497. The procedure involves immunizing a host animal (typically a mouse because of the availability of murine myelomas) with the protein of interest.

Immortal antibody-producing cell lines can be created by cell fusion, and also by other techniques such as direct transformation of B lymphocytes with oncogenic DNA, or transfection with Epstein-Barr virus. *See, e.g.*, M. Schreier *et al.*,

Hybridoma Techniques (1980); Hammerling *et al.*, *Monoclonal Antibodies and T-cell Hybridomas* (1981); Kennett *et al.*, *Monoclonal Antibodies* (1980); *see*, also, U.S. Patent Nos. 4,341,761; 4,399,121; 4,427,783; 4,444,887; 4,452,570; 4,466,917; 4,472,500, 4,491,632; and 4,493,890. Monoclonal antibodies are useful for purifying the individual antigens which they are directed against. Furthermore, both polyclonal and monoclonal antibodies can be used in vaccine compositions to impart passive immunization to an individual to which the compositions are administered.

The present system has been exemplified herein with respect to hypervariable regions present in the gp120 envelope protein of HIV-1. Two basic types of HIV, known as "HIV-1" and "HIV-2," respectively, have currently been identified. Additionally, at least five genetic subtypes of HIV-1 (termed subtypes "A-E") have been characterized, based on phylogenetic relationships determined by using the HIV *gag* and *env* genes. Each of these subgroups is made up of numerous viral isolates. For example, subtype B includes the North American/European isolates such as HIV_{IIIb}, HIV_{SF2}, HIV_{LAV}, HIV_{LAI}, HIV_{MN}, etc. Exemplified herein are constructs derived from viral isolates from two of the five subtypes. In particular, gp120 from the North American/European SF2 HIV-1 isolate, found in subtype B, and a Thailand isolate, termed, N. Thai CM235, from subtype E, have been used herein as backbone proteins to illustrate the invention. However, the gp120 sequence for a multitude of HIV-1 and HIV-2 isolates is known and reported (*see*, *e.g.*, Myers *et al.* Los Alamos Database, Los Alamos National Laboratory, Los Alamos, New Mexico (1992); Myers *et al.*, *Human Retroviruses and Aids, 1990*, Los Alamos, New Mexico: Los Alamos National Laboratory; and Modrow *et al. J. Virol.* (1987) 61:570-578, for a comparison of the envelope gene sequences of a variety of HIV isolates) and the basic gp120 backbone can be derived from any of these various seqeunces, depending on the viral isolate or group of isolates to be targeted.

At least five hypervariable domains have been identified in gp120 (termed "V1-V5"). These domains occur at approximately amino acid positions 125-155 (V1), 155-199 (V2), 299-333 (V3), 387-415 (V4) and about 457-469 (V5), numbered according to the gp120 sequence of HIV-1 SF2, depicted in Figure 1. The regions are bounded by cysteines, except the V5 region. The hypervariable domains

are characterized as lacking substantial homology (*e.g.*, as low as 10% homology) in differing HIV isolates, particularly from differing subtypes. Furthermore, there is substantial variation in length among the hypervariable regions from various isolates due to the prevalence of insertion and deletion mutations. Thus, these regions lack any substantial degree of amino acid sequence homology, and can only be assigned an approximate length.

The primary characterization of the hypervariable domains is their location within the envelope glycoprotein and their presumed tertiary structure (*i.e.*, loops due to the presence of cysteines). The conserved or constant domains, as well as all 18 cysteines in the envelope, are highly conserved. Corresponding hypervariable domains are located in identical positions relative to both surrounding constant domains and cysteines, from one isolate to the next. Furthermore, the tertiary structure of the hypervariable domains appears to be highly conserved, *e.g.*, two non-homologous hypervariable domains from different HIV isolates will usually both exhibit the same three- dimensional conformation, such as an exposed loop. Thus, hypervariable regions from new HIV isolates can be readily identified by sequencing the new isolates and comparing the sequence to known HIV sequences so that the conserved domains and cysteines are aligned. Accordingly, hypervariable domains from newly identified isolates will also find use with the present invention.

The V3 region has been extensively characterized and V3 consensus sequences for all five of the HIV-1 subtypes determined (see Figures 2A-E). The consensus sequences were determined by comparing numerous viral isolates and assigning the most common amino acid encountered at a particular site to that position. However, it is to be understood that these consensus sequences may evolve as new sequences from each subtype become available. Hence, the V3 consensus sequences for use in the present invention will include not only those depicted, but newly determined sequences as more isolates are characterized.

As explained above, the V3 region spans approximately amino acid positions 299-333, inclusive, of Figure 1. However, it appears that less than the full V3 region is essential for immunoreactivity. In particular, the amino acid sequences spanning positions 308-323, and more particularly 313-318 (Page *et al. J. Virol.* (1992)

66:524-533) appear to be important regions for immunogenicity. Accordingly, the constructs can include consensus sequences derived from this core region and need not include the sequence for the full V3 domain, so long as an active gp120 (as determined, *e.g.*, by the CD4 binding assays described below) is produced.

The consensus sequences for subtypes B and E (occurring on the top line of Figures 2B and 2E, respectively) served as models for developing the present system. However, consensus sequences derived from any of the other subtypes, as well as those derived from any of the other hypervariable regions described above which show immunoreactivity, may also be used in the present constructs. As explained above, the consensus sequences are flanked with unique restriction sites such that an existing sequence can easily be excised from an expression vector and other consensus sequences inserted.

The HIV gp120 constructs of the present invention can be conveniently expressed in eucaryotic or procaryotic cell systems, described above, but are preferably prepared in mammalian expression systems. The protein can be purified using standard techniques. One purification technique for the recombinant gp120 is the method described in International Publication No. WO 91/13906 (published 19 September 1991). Alternatively, the proteins can be purified using antibodies, either polyclonal or monoclonal, directed against gp120. *See*, *e.g.*, International Publication No. WO91/15238 (published 17 October 1991). Activity can be determined using a CD4 binding assay which employs standard radioimmune precipitation (RIP) techniques. For the RIP procedure, anti-gp120 antibodies can be bound to a suitable substrate, such as Protein A Sepharose, ^{35}S-labeled CD4 which has been preincubated with the test sample, can then be added. After immunoprecipitation, the samples can be solubilized, boiled and applied to gels. Other suitable assays for the detection of CD4 binding are known and include, for example, gel filtration HPLC.

As explained above, the technique is applicable to other RNA viruses possessing hypervariable regions. Examples of such viruses include Hog Cholera Virus, Bovine Viral Diarrhea Virus (BVDV), Hepatitis C Virus (HCV), the Dengue and Yellow Fever Virus, the influenza viruses among others. In particular, at least six-seven groups of HCV have been identified based on sequence homology at the amino

acid and nucleotide level. *See*, Houghton *et al. Hepatolgy* (1991) 14:381-388. Thus, as with HIV, consensus sequences for immunogenic hypervariable regions in each of the groups can be determined and vectors constructed with unique restriction sites, allowing for easy insertion and excision of a particular consensus sequence from any of the various groups of HCV.

The viral genomic sequence of HCV is known, as are methods for obtaining the sequence. *See*, *e.g.*, International Publication Nos. WO89/04669; WO90/11089; and WO90/14436. The genome encodes at least two envelope proteins termed E1 and E2/NS1, respectively. HCV-1 E2/NS1 includes an N-terminal hypervariable region of approximately 30 amino acids, occurring at positions 384-414 of the protein, (numbered according to the HCV-1 sequence, see Figure 3 of WO93/06126, published 1 April 1993) and particularly amino acid positions 396-407. See Figure 6 for a comparison of the HCV E2/NS1 hypervariable regions for 90 HCV isolates and the determined consensus sequence based on these sequences. Similarities between this hypervariable region and HIV-1 gp120 V3 exist with respect to the degree of sequence variation, the predictive effect of amino acid changes on putative antibody binding, as well as the lack of defined secondary structure. This region has been shown to be immunogenic through antibody epitope mapping experiments described in International Publication No. WO93/06126 (published 1 April 1993).

Similarly, an E1 variable domain found within amino acids 215-255 (numbered according to the HCV-1 sequence, see Figure 2 of WO93/06126, published 1 April 1993) is also present in the genome. As with the E2/NS1 hypervariable region, this domain appears to be an important immunoreactive domain.

Accordingly, consensus sequences can be determined for these regions in a given group of the virus and, as with gp120 V3, the consensus sequence can be used to replace all or part of the wild-type hypervariable region in genetic constructs to provide immunogenic proteins for the diagnosis and protection against a wide array of viral isolates. Again, the sequences are conveniently flanked with unique restriction sites so that a consensus sequence determined for one group can be easily replaced in the construct with a consensus sequence determined for another of the HCV groups.

The HCV constructs can be expressed in both eucaryotic and procaryotic expression systems, as described above, and purified using known techniques. Standard assays such as ELISAs, RIPs, Western blots etc. can be used to determine activity of the engineered proteins.

Influenza virus is another example of an RNA virus for which the present invention will be particularly useful. Specifically, the envelope glycoproteins HA and NA, of influenza A, show considerable variability from isolate to isolate. Numerous HA subtypes of influenza A have been identified (Kawaoka et al. Virology (1990) 179:759-767; Webster et al. "Antigenic variation among type A influenza viruses," p. 127-168. In: P. Palese and D.W. Kingsbury (ed.), *Genetics of influenza viruses*. Springer-Verlag, New York), based on sequence heterogeneity. HA from subtype H1 includes a variable region known as antigenic site Sa and subtype H3 includes a corresponding region termed antigenic site B. Site B contains an exposed loop (found at amino acid positions 155-160 in the H3 numbering system), shows considerable variability and is highly immunogenic. Accordingly, this region is of particular interest and can be engineered, based on the isolates in question, and used in the present constructs.

The activity of the expressed hemagglutinins can be tested using standard viral and hemagglutinin inhibition assays. For example, the ability of the above-engineered hemagglutinin to inhibit influenza virus adsoption to erythrocytes can be examined via the method of Pritchett *et al. Virology* (1987) 160:502-506. In this assay, virus is incubated with human erythrocytes in the presence and absence of the protein in question. The bound virus is then quantitated by measuring the amount of viral sialidase associated with the cells after centrifugation and washing. Other methods of testing for inhibition are also known and will readily find use herein. See, *e.g.*, Palmer *et al. Immunol. Ser.* (1975) 6:51-52; Li *et al. J. Virol.* (1992) 66:399-404.

The polypeptides expressed using the present system and/or antibodies generated against the same, can be used as diagnostics to detect the presence of reactive antibodies and/or antigens of the viral isolate or isolates of interest, in a biological sample. For example, the presence of antibodies reactive with the engineered proteins

and, conversely, antigens reactive with antibodies generated thereto, can be detected using standard electrophoretic and immunodiagnostic techniques, including immunoassays such as competition, direct reaction, or sandwich type assays. Such assays include, but are not limited to, Western blots; agglutination tests; enzyme-labeled and mediated immunoassays, such as ELISAs; biotin/avidin type assays; radioimmunoassays; immunoelectrophoresis; immunoprecipitation, etc. The reactions generally include revealing labels such as fluorescent, chemiluminescent, radioactive, or enzymatic labels or dye molecules, or other methods for detecting the formation of a complex between the antigen and the antibody or antibodies reacted therewith.

Solid supports can be used in the assays such as nitrocellulose, in membrane or microtiter well form; polyvinylchloride, in sheets or microtiter wells; polystyrene latex, in beads or microtiter plates; polyvinylidine fluoride; diazotized paper; nylon membranes; activated beads, and the like. Typically, the solid support is first reacted with the biological sample (or engineered proteins), washed and then the antibodies, (or a sample suspected of containing antibodies), applied. If a sandwich type format is desired, such as a sandwich ELISA assay, a commercially available anti-immunoglobulin (i.e. anti-rabbit immunoglobulin) conjugated to a detectable label, such as horseradish peroxidase, alkaline phosphatase or urease, can be added. An appropriate substrate is then used to develop a color reaction.

Alternatively, a "two antibody sandwich" assay can be used to detect the proteins of the present invention. In this technique, the solid support is reacted first with one or more of the antibodies of the present invention, washed and then exposed to the test sample. Antibodies are again added and the reaction visualized using either a direct color reaction or using a labeled second antibody, such as an anti-immunoglobulin labeled with horseradish peroxidase, alkaline phosphatase or urease.

Assays can also be conducted in solution, such that the viral proteins and antibodies thereto form complexes under precipitating conditions. The precipitated complexes can then be separated from the test sample, for example, by centrifugation. The reaction mixture can be analyzed to determine the presence or absence of antibody-antigen complexes using any of a number of standard methods, such as those immunodiagnostic methods described above.

The above-described antigens and antibodies can be provided in kits, with suitable instructions and other necessary reagents, in order to conduct immunoassays as described above. The kit can also contain, depending on the particular immunoassay used, suitable labels and other packaged reagents and materials (i.e. wash buffers and the like). Standard immunoassays, such as those described above, can be conducted using these kits.

The recombinant antigens or antibodies generated therefrom, can also be formulated into vaccine compositions to provide immunity to a broad spectrum of viral isolates. These vaccines may either be prophylactic (to prevent infection) or therapeutic (to treat disease after infection). Additionally, the vaccines can comprise mixtures of one or more of the engineered proteins, such as a gp120 backbone having a subtype B consensus sequence, a gp120 backbone with a subtype E consensus sequence, a gp120 backbone with a subtype A consensus sequence, and so on.

These vaccines can include one or more "pharmaceutically acceptable excipients or vehicles" such as water, saline, glycerol, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles.

A carrier is optionally present which is a molecule that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen may be conjugated to a bacterial toxoid, such as toxoid from diphtheria, tetanus, cholera, etc.

Adjuvants may also be used to enhance the effectiveness of the vaccines. Such adjuvants include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc.; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example

(a) MF59 (PCT Publication No. WO90/14837), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particle generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freunds Adjuvant (CFA) and Incomplete Freunds Adjuvant (IFA); (5) cytokines, such as interleukins (IL-1, IL-2, etc.), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc.; and (6) other substances that act as immunostimulating agents to enhance the effectiveness of the composition. Alum and MF59 are preferred.

Muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acteyl-normuramyl-L-alanyl-D-isogluatme (nor-MDP), N-acetylmuramyl-L-alanyl-D-isogluatminyl-L-alanine-2-(1'-2'-dipalmitoyl-*sn*-glycero-3-huydroxyphosphoryloxy)-ethylamine (MTP-PE), etc.

Typically, the immunogenic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation also may be emulsified or encapsulated in liposomes for enhanced adjuvant effect, as discussed above.

Immunogenic compositions used as vaccines comprise a therapeutically effective amount of the antigenic polypeptides, as well as any other of the above-mentioned components, as needed. By "therapeutically effective amount" is meant an amount of epitope-bearing polypeptide sufficient to induce an immunological response (as defined above) in the individual to which it is administered. Preferably, the

effective amount is sufficient to bring about treatment, also as defined above. The exact amount necessary will vary depending on the intended use of the polypeptide. For example, if the polypeptide is to be used in vaccine compositions or for the generation of polyclonal antiserum or antibodies, the effective amount will depend on the taxonomic group of individual to be treated (*e.g.*, nonhuman primate, primate, etc.); age and general condition of the individual to be treated; the capacity of the individual's immune system to synthesize antibodies; the degree of protection desired; the severity of the condition being treated; the particular polypeptide selected and its mode of administration, among other factors. An appropriate effective amount can be readily determined by one of skill in the art. A "therapeutically effective amount" will fall in a relatively broad range that can be determined through routine trials.

The immunogenic compositions are conventionally administered parenterally, e.g., by injection, either subcutaneously or intramuscularly. Additional formulations suitable for other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. The vaccine may be administered in conjunction with other immunoregulatory agents.

In addition to the above, it is also possible to prepare live vaccines of attenuated microoorganisms which express recombinant polypeptides from the selected vectors. Suitable attenuated microorganisms are known in the art and include for example, viruses (*e.g.*, *Vaccinia*, and fowl pox virus), and bacteria (*e.g.*, cholera, *Salmonella*, *Bacille Clamette-Guerin* (tuberculosis), *Helicobacter pylori*, etc.).

In addition, the vaccine may be administered in conjunction with other immunoregulatory agents, for example, immune globulins.

III. Experimental

Below are examples of specific embodiments for carrying out the present invention. The examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way.

Efforts have been made to ensure accuracy with respect to numbers used (*e.g.*, amounts, temperatures, etc.), but some experimental error and deviation should, of course, be allowed for.

Example 1

Construction of plasmid pCMVgp120$_{SF2.3/NAE.C}$

Containing an HIV-1 Subtype B VP3 Consensus Sequence in a Subtype B gp120 Backbone

A novel gp120, including the subtype B V3 consensus sequence in a gp120 backbone derived from HIV-1 SF2, flanked by unique restriction sites, was constructed as follows. The wild-type gp120 backbone from the HIV-1 isolate, SF2, is shown in Figure 1. (It should be noted that thr-30 in the SF2 wild-type gp120 has been substituted conservatively with a ser in constructs used in the present invention.) The V3 region (underlined) spans positions cys-299 through cys-333, of the figure. Positions flanking the V3 region where unique restriction sites could be designed were determined. Specifically identified were potential 5' *Nru*I and 3' *Xba*I sites (Figure 4). *Bgl*II and *Bsu*36I sites were also identified which were further upstream and downstream, respectively, of the *Nru*I and *Xba*I sites (Figure 4).

Overlapping synthetic oligonucleotides corresponding to the consensus V3 region, the unique restriction sites *Nru*I and *Xba*I, and DNA that restored the sequences to the flanking cloning sites (*Bgl*II and *Bsu*36I), were prepared. The V3 loop synthetic oligonucleotide included sequence changes at nucleotide positions 931, 955, 967, 968, and 988, to code for the amino acids his, tyr, glu and gln, at positions 311, 319, 323, and 330, respectively, instead of the wild-type sequence found at these positions (Figures 1 and 3). Similarly, the sequences at the potential *Nru*I and *Xba*I sites were manipulated to ensure cleavage by the restriction enzymes at these sites. Specifically, nucleotide positions 885 and 888 were changed to C and G, from A and A, respectively, to generate the NruI site, and nucleotide positions 1006 and 1007 were

changed to T and C from A and G, respectively, to generate the XbaI site (Figures 1 and 3).

The engineered sequence was inserted in place of the wild-type sequence into plasmid pCMV6a120-SF2 (ATCC Accession No. 68249, Figure 5) rendering plasmid $pCMVgp120_{SF2.3/NAE.C}$ (ATCC Accession No. 39365). This plasmid bears the SF2 sequence for gp120 with the genetically engineered subtype B consensus V3 loop, the tPA signal sequence, the SV40 early promoter and enhancer, the SV40 polyadenylation site, and an SV40 origin of replication for use of the vector in mammalian cells.

Plasmid $pCMVgp120_{SF2.3/NAE.C}$ was used to transform COS7 cells. These cells were used because they supply SV40 T-Ag in trans which allows for the gp120 DNA templates to replicate to a high copy number (approximately 10,000/cell). Immunoprecipitation of radiolabeled extracts from consensus transfected COS7 cells demonstrated the synthesis of an immunoreactive gp120 molecule with mobility similar to parental SF2 gp120.

Example 2

Construction of plasmid $pCMVgp120_{SF2.3/NT9142}$ Containing an HIV-1 Subtype E VP3 Consensus Sequence in a Subtype B gp120 Backbone

The subtype B consensus sequence, present in plasmid $pCMVgp120_{SF2.3/NAE.C}$, described above, was replaced with a subtype E V3 consensus sequence, derived from the N. Thai isolate, 9142, as follows (Figures 1 and 3). Similarly, the sequences at the potential *Nru*I and *Xba*I sites were manipulated to ensure cleavage by the restriction enzymes at these sites. Specifically, nucleotide positions 885 and 888 were changed to C and G, from A and A, respectively, to generate the NruI site, and nucleotide positions 1006 and 1007 were changed to T and C from A and G, respectively, to generate the XbaI site (Figures 1 and 3), rendering plasmid $pCMVgp120_{SF2.3/NT9142}$ (ATCC Accession No. 69366). The plasmid was used to transform COS7 cells, as above. The expressed protein showed immunoreactivity

using a standard radio immunoprecipitation (RIP) assay, using a Thailand-derived serum from a patient infected with HIV.

Example 3

Construction of plasmid pCMVgp120$_{CM235.3/NT9142}$ Containing an HIV-1 Subtype E VP3 Consensus Sequence in a Subtype E gp120 Backbone

gp120, from clone CM235 served as the backbone for this construction. The clone was derived from an N. Thai isolate from subtype E. The subtype E V3 consensus sequence (Figure 2) was engineered into this backbone as follows. An overlapping PCR product was created using three PCR reactions and CM235 as the template: 1) a 3' primer complementary to part of the V3 sequence, also including nucleotides which overlapped the V3 region with the desired changes; 2) a 5' primer complementary to part of the V3 sequence including nucleotides which overlapped the V3 region with the desired changes; and 3) end primers. The resulting PCR products were reamplified with end primers, and the final PCR product encoded the subtype E consensus sequence in the CM235 backbone, having Thr instead of Pro at position 13 of V3, and Val instead of Ala at position 19 of V3 (see Figure 2E). The engineered sequence was inserted in place of the wild-type sequence into plasmid pCMV6a120-SF2 (described above) rendering plasmid pCMVgp120$_{CM235.3/NT9142}$ (ATCC Accession No. 69367).

The plasmid was used to transform COS7 cells, as above. The expressed protein showed immunoreactivity using a standard radioimmunoprecipitation assay with serum from a Thai patient infected with HIV.

Example 4

CD4 Binding Assays

The ability of the V3 consensus sequence-containing gp120 to bind CD4, thus consistent with the polypeptide being in native conformation, was determined using a radioimmune precipitation (RIP) assay. The CD4 used in the assay was a recombinant, soluble CD4 derived from a CHO cell line transfected with an expression plasmid encoding the full external domain, termed "CHO ST4.2" (ERC BioServices Corporation, Rockville, Md). CD4 was labeled as follows. Confluent monolayers of

the CHO ST4.2 cells were labeled in Dulbecco's modified Eagle medium (DME) without cysteine and methionine (cys-met-DME). Five ml of cys-met-DME with 500 μCi each ^{35}S-met and ^{35}S-cys (New England Nuclear), were added and the cells were labeled for 4.5 hours. The cells were harvested, centrifuged and the supernatant stored frozen for future use.

Samples for use in the RIPs were prepared as follows. COS7 cells were washed with DME and $CaPO_4$ transfected with (1) plasmid pCMV6a120-SF2, including the gene for wild-type SF2 gp120; (2) plasmid pCMVgp120$_{SF2.3/NAE.C}$, described above (or other plasmids of the present invention), containing the subtype B V3 consensus sequence in a subtype B gp120 backbone; or (3) pGEM3Z (Promega), a cloning vector which does not contain any eukaryotic coding sequences. Transfected cells were incubated for three hours, subjected to glycerol shock for one hour, incubated for 48 hours and 10 mls media removed and frozen. 5 mls more media were added for an additional 24 hours, after which time the media was harvested and pooled. The samples were centrifuged and the supernatant concentrated three times using an Amicon 30 filter.

The samples from above were assayed as follows. One ml of the media from above, with 100 μl 10x lysis buffer (LB, 100mM NaCl, 20mM Tris 7.5, 1mM EDTA, 0.5%NP40 and 0.5% deoxycholate), was incubated with ^{35}S-CD4, labeled as described above. Protein A Sepharose was washed and 25 μl of a 1:200 dilution of goat-anti gp120, was added. The antibody was absorbed to the Sepharose for 2.5 hours and the substrate was washed two times. The labeled media from above was added and immune precipitations allowed to proceed in the cold. The reaction was terminated by washing with LB. The samples were solubilized in Laemmli buffer, boiled and applied to 11.5% polyacrylamide gels. Gels were treated with En^{3}Hance®, dried, exposed and developed. The results of the assay demonstrated that the consensus sequence-containing gp120 bound CD4, indicating that the molecule was conformationally correct.

Thus, novel constructs for the expression of immunogenic viral proteins containing consensus sequences in place of corresponding hypervariable regions have been disclosed. Although preferred embodiments of the subject invention have been

described in some detail, it is understood that obvious variations can be made without departing from the spirit and the scope of the invention as defined by the appended claims.

Deposits of Strains Useful in Practicing the Invention

A deposit of biologically pure cultures of the following strains was made with the American Type Culture Collection (ATCC), 12301 Parklawn Drive, Rockville, Maryland. The accession number indicated was assigned after successful viability testing, and the requisite fees were paid. Access to said cultures will be available during pendency of the patent application to one determined by the Commissioner to be entitled thereto under 37 CFR 1.14 and 35 USC 122. All restriction on availability of said cultures to the public will be irrevocably removed upon the granting of a patent based upon the application. Moreover, the designated deposits will be maintained for a period of thirty (30) years from the date of deposit, or for five (5) years after the last request for the deposit; or for the enforceable life of the U.S. patent, whichever is longer. Should a culture become nonviable or be inadvertently destroyed, or, in the case of plasmid-containing strains, lose its plasmid, it will be replaced with a viable culture(s) of the same taxonomic description.

These deposits are provided merely as convenience to those of skill in the art, and are not an admission that a deposit is required under 35 USC §112. The nucleic acid sequences of these plasmids, as well as the amino acid sequences of the polypeptides encoded thereby, are incorporated herein by reference and are controlling in the event of any conflict with the description herein. A license may be required to make, use, or sell the deposited materials, and no such license is hereby granted.

Strain	Deposit Date	ATCC No.
pCMV6a120-SF2 in *E. coli* HB101	March 8, 199	068249
$pCMVgp120_{SF2.3/NAE.C}$ in *E. coli* HB101	July 27, 199	369365
$pCMVgp120_{SF2.3/NT9142}$ in *E. coli* HB101	July 27, 199	369366
$pCMVgp120_{CM235.3/NT9142}$ in *E. coli* HB101	July 27, 199	369367

CLAIMS

1. A recombinant construct comprising a nucleotide sequence encoding a backbone immunogenic protein derived from a parental RNA virus, said backbone protein characterized as having at least one wild-type hypervariable region in its native state, wherein a sequence from said wild-type hypervariable region is substituted with a replacement sequence corresponding thereto, and further wherein said replacement sequence is flanked at the 3'- and 5'-ends thereof with unique restriction sites.

2. The recombinant construct of claim 1 further comprising control elements that are operably linked to said replacement sequence whereby said nucleotide sequence can be transcribed and translated in a host cell and at least one of said control elements is heterologous to said nucleotide sequence.

3. The recombinant construct of claim 1 wherein said replacement sequence comprises a consensus sequence determined for a selected group of viral isolates of said parental RNA virus.

4. The recombinant construct of claim 1 wherein said replacement sequence comprises a hypervariable region from a viral isolate related to said parental RNA virus.

5. The recombinant construct of claim 1 wherein said RNA virus is a retrovirus.

6. The recombinant construct of claim 5 wherein said retrovirus is a human immunodeficiency virus (HIV).

7. The recombinant construct of claim 6 wherein said HIV is HIV-1.

8. The recombinant construct of claim 7 wherein said hypervariable region is the V3 region of HIV-1 gp120.

9. The recombinant construct of claim 8 wherein said replacement sequence is selected from the group consisting of HIV-1 subtype A V3, subtype B V3, subtype C V3, subtype D V3, and subtype E V3.

10. The recombinant construct of claim 9 wherein said replacement sequence comprises the consensus sequence for HIV-1 subtype B V3.

11. The recombinant construct of claim 9 wherein said replacement sequence comprises the consensus sequence for HIV-1 subtype E V3.

12. A recombinant vector comprising:

(a) a DNA sequence encoding a human immunodeficiency virus type 1 (HIV-1) subtype B gp120 wherein the wild-type V3 hypervariable region is replaced with a DNA sequence encoding a corresponding consensus sequence comprising the sequence CTRPNNNTRKSIHIGPGRAFYTTGEIIGDIRQAHC; and

(b) control elements that are operably linked to said DNA sequence whereby said DNA sequence can be transcribed and translated in a host cell and at least one of said control elements is heterologous to said DNA sequence.

13. A recombinant vector comprising:

(a) a DNA sequence encoding a human immunodeficiency virus type 1 (HIV-1) subtype E gp120 wherein the wild-type V3 hypervariable region is replaced with a DNA sequence encoding a corresponding consensus sequence comprising the sequence CTRPSNNTRTSITIGPGQVFYRTGDIIGDIRKAYC; and

(b) control elements that are operably linked to said DNA sequence whereby said DNA sequence can be transcribed and translated in a host cell and at least one of said control elements is heterologous to said nucleotide sequence.

14. Plasmid pCMVgp120$_{SF2.3/NAE.C}$ (ATCC No. 69365).

15. Plasmid pCMVgp120$_{SF2.3/NT9142}$ (ATCC No. 69366).

16. Plasmid pCMVgp120$_{CM235.3/NT9142}$ (ATCC No. 69367).

17. A host cell transformed with the recombinant construct of claim 2.

18. The host cell of claim 17 wherein the host cell is a mammalian cell.

19. A host cell transformed with the plasmid of claim 14.

20. A host cell transformed with the plasmid of claim 15.

21. A host cell transformed with the plasmid of claim 16.

22. A method of producing a recombinant polypeptide comprising:

(a) providing a population of host cells according to claim 17; and

(b) culturing said population of cells under conditions whereby the polypeptide encoded by said nucleotide sequence is expressed.

23. A method of producing a recombinant polypeptide comprising:

(a) providing a population of host cells according to claim 18; and

(b) culturing said population of cells under conditions whereby the gp120 encoded by said plasmid is expressed.

24. A method of producing a recombinant polypeptide comprising:

(a) providing a population of host cells according to claim 19; and

(b) culturing said population of cells under conditions whereby the gp120 encoded by said plasmid is expressed.

25. A method of producing a recombinant polypeptide comprising:

(a) providing a population of host cells according to claim 20; and

(b) culturing said population of cells under conditions whereby the gp120 encoded by said plasmid is expressed.

26. A method of producing a recombinant polypeptide comprising:

(a) providing a population of host cells according to claim 21; and

(b) culturing said population of cells under conditions whereby the gp120 encoded by said plasmid is expressed.

27. A human immunodeficiency virus (HIV) gp120 wherein the wild-type V3 sequence is replaced in whole or part with a consensus sequence corresponding thereto.

28. The gp120 of claim 27 which is derived from HIV-1.

29. The gp120 of claim 28 which is derived from the group consisting of HIV-1 subtype A, subtype B, subtype C, subtype D, and subtype E.

30. The gp120 of claim 29 which is derived from HIV-1 subtype B.

31. The gp120 of claim 28 which is derived from HIV-1 subtype E.

32. The gp120 of claim 30 wherein said consensus sequence comprises the amino acid sequence CTRPNNNTRKSIHIGPGRAFYTTGEIIGDIRQAHC.

33. A vaccine composition comprising the gp120 of claim 27 in combination with a pharmaceutically acceptable excipient.

34. The vaccine composition of claim 33 further comprising an adjuvant.

35. The vaccine composition of claim 34 wherein said adjuvant is MF59.

36. A method of making a vaccine composition comprising admixing a gp120 according to claim 27 with a pharmaceutically acceptable excipient.

37. A method of treating or preventing human immunodeficiency virus infection in an individual comprising administering a therapeutically effective amount of the vaccine composition of claim 33 to said individual.

```
1    ATG AAA GTG AAG GGG ACC AGG AGG AAT TAT CAG CAC TTG TGG AGA TGG GGC ACC TTG CTC
1    met lys val lys gly thr arg arg asn tyr gln his leu trp arg trp gly thr leu leu
                                                  → mature gp120
61   CTT GGG ATG TTG ATG ATC TGT AGT GCT ACA GAA AAA TTG TGG GTC ACA GTT TAT TAT GGA
21   leu gly met leu met ile cys ser ala thr glu lys leu trp val thr val tyr tyr gly

121  GTA CCT GTG TGG AAA GAA GCA ACT ACC ACT CTA TTT TGT GCA TCA GAT GCT AGA GCA TAT
41   val pro val trp lys glu ala thr thr thr leu phe cys ala ser asp ala arg ala tyr

181  GAT ACA GAG GTA CAT AAT GTT TGG GCC ACA CAT GCC TGT GTA CCC ACA GAC CCC AAC CCA
61   asp thr glu val his asn val trp ala thr his ala cys val pro thr asp pro asn pro

241  CAA GAA GTA GTA TTG GGA AAT GTG ACA GAA AAT TTT AAC ATG TGG AAA AAT AAC ATG GTA
81   gln glu val val leu gly asn val thr glu asn phe asn met trp lys asn asn met val

301  GAA CAG ATG CAG GAG GAT ATA ATC AGT TTA TGG GAT CAA AGC CTA AAG CCA TGT GTA AAA
101  glu gln met gln glu asp ile ile ser leu trp asp gln ser leu lys pro cys val lys

361  TTA ACC CCA CTC TGT GTT ACT TTA AAT TGC ACT GAT TTG GGG AAG GCT ACT AAT ACC AAT
121  leu thr pro leu cys val thr leu asn cys thr asp leu gly lys ala thr asn thr asn

421  AGT AGT AAT TGG AAA GAA GAA ATA AAA GGA GAA ATA AAA AAC TGC TCT TTC AAT ATC ACC
141  ser ser asn trp lys glu glu ile lys gly glu ile lys asn cys ser phe asn ile thr

481  ACA AGC ATA AGA GAT AAG ATT CAG AAA GAA AAT GCA CTT TTT CGT AAC CTT GAT GTA GTA
161  thr ser ile arg asp lys ile gln lys glu asn ala leu phe arg asn leu asp val val

541  CCA ATA GAT AAT GCT AGT ACT ACT ACC AAC TAT ACC AAC TAT AGG TTG ATA CAT TGT AAC
181  pro ile asp asn ala ser thr thr thr asn tyr thr asn tyr arg leu ile his cys asn

601  AGA TCA GTC ATT ACA CAG GCC TGT CCA AAG GTA TCA TTT GAG CCA ATT CCC ATA CAT TAT
201  arg ser val ile thr gln ala cys pro lys val ser phe glu pro ile pro ile his tyr

661  TGT ACC CCG GCT GGT TTT GCG ATT CTA AAG TGT AAT AAT AAA ACG TTC AAT GGA AAA GGA
221  cys thr pro ala gly phe ala ile leu lys cys asn asn lys thr phe asn gly lys gly

721  CCA TGT ACA AAT GTC AGC ACA GTA CAA TGT ACA CAT GGA ATT AGG CCA ATA GTG TCA ACT
241  pro cys thr asn val ser thr val gln cys thr his gly ile arg pro ile val ser thr

781  CAA CTG CTG TTA AAT GGC AGT CTA GCA GAA GAA GAG GTA GTA ATT AGA TCT GAC AAT TTC
261  gln leu leu leu asn gly ser leu ala glu glu glu val val ile arg ser asp asn phe

841  ACG AAC AAT GCT AAA ACC ATA ATA GTA CAG CTG AAT GAA TCT GTA GCA ATT AAC TGT ACA
281  thr asn asn ala lys thr ile ile val gln leu asn glu ser val ala ile asn cys thr

901  AGA CCC AAC AAC AAT ACA AGA AAA AGT ATC TAT ATA GGA CCA GGG AGA GCA TTT CAT ACA
301  arg pro asn asn asn thr arg lys ser ile tyr ile gly pro gly arg ala phe his thr

961  ACA GGA AGA ATA ATA GGA GAT ATA AGA AAA GCA CAT TGT AAC ATT AGT AGA GCA CAA TGG
321  thr gly arg ile ile gly asp ile arg lys ala his cys asn ile ser arg ala gln trp

1021 AAT AAC ACT TTA GAA CAG ATA GTT AAA AAA TTA AGA GAA CAG TTT GGG AAT AAT AAA ACA
341  asn asn thr leu glu gln ile val lys lys leu arg glu gln phe gly asn asn lys thr

1081 ATA GTC TTT AAT CAA TCC TCA GGA GGG GAC CCA GAA ATT GTA ATG CAC AGT TTT AAT TGT
361  ile val phe asn gln ser ser gly gly asp pro glu ile val met his ser phe asn cys

1141 AGA GGG GAA TTT TTC TAC TGT AAT ACA ACA CAA CTG TTT AAT AAT ACA TGG AGG TTA AAT
381  arg gly glu phe phe tyr cys asn thr thr gln leu phe asn asn thr trp arg leu asn

1201 CAC ACT GAA GGA ACT AAA GGA AAT GAC ACA ATC ATA CTC CCA TGT AGA ATA AAA CAA ATT
401  his thr glu gly thr lys gly asn asp thr ile ile leu pro cys arg ile lys gln ile
```

Figure 1A

```
1261 ATA AAC ATG TGG CAG GAA GTA GGA AAA GCA ATG TAT GCC CCT CCC ATT GGA GGA CAA ATT
 421 ile asn met trp gln glu val gly lys ala met tyr ala pro pro ile gly gly gln ile

1321 AGT TGT TCA TCA AAT ATT ACA GGG CTG CTA TTA ACA AGA GAT GGT GGT ACA AAT GTA ACT
 441 ser cys ser ser asn ile thr gly leu leu leu thr arg asp gly gly thr asn val thr

1381 AAT GAC ACC GAG GTC TTC AGA CCT GGA GGA GGA GAT ATG AGG GAC AAT TGG AGA AGT GAA
 461 asn asp thr glu val phe arg pro gly gly gly asp met arg asp asn trp arg ser glu

1441 TTA TAT AAA TAT AAA GTA ATA AAA ATT GAA CCA TTA GGA ATA GCA CCC ACC AAG GCA AAG
 481 leu tyr lys tyr lys val ile lys ile glu pro leu gly ile ala pro thr lys ala lys

1501 AGA AGA GTG GTG CAG AGA GAA AAA AGA TGA TGA
 501 arg arg val val gln arg glu lys arg OPA OPA
                  mature gp120
```

Figure 1B

841 ACG AAC AAT GCT AAA ACC ATA ATA GTA CAG CTG AAT GAA TCT GTC GGG ATT AAC TGT ACA
281 thr asn asn ala lys thr ile ile val gln leu asn glu ser val ala ile asn cys thr

901 AGA CCC AAC AAC AAT ACA AGA AAA AGT ATC CAT ATA GGA CCA GGG AGA GCA TTT TAT ACA
301 arg pro asn asn asn thr arg lys ser ile his ile gly pro gly arg ala phe tyr thr

961 ACA GGA GAA ATA ATA GGA GAT ATA AGA CAA GCA CAT TGT AAC ATT TCT AGA GCA CAA TGG
321 thr gly glu ile ile gly asp ile arg gln ala his cys asn ile ser arg ala gln trp

FIGURE 2

CTRPNNNTRKSVHIGPGQAFYATGDIIGDIRQAHC

2A

CTRPNNNTRKSIHIGPGRAFYTTGEIIGDIRQAHC

2B

CTRPNNNTRKSIRIGPGQTFYATGDIIGDIRQAHC

2C

CTRPYNNTRQRTHIGPGQALYTT-RIIGDIRQAHC

2D

CTRPSNNTRTSITIGPGQVFYRTGDIIGDIRKAYC

2E

FIGURE 3

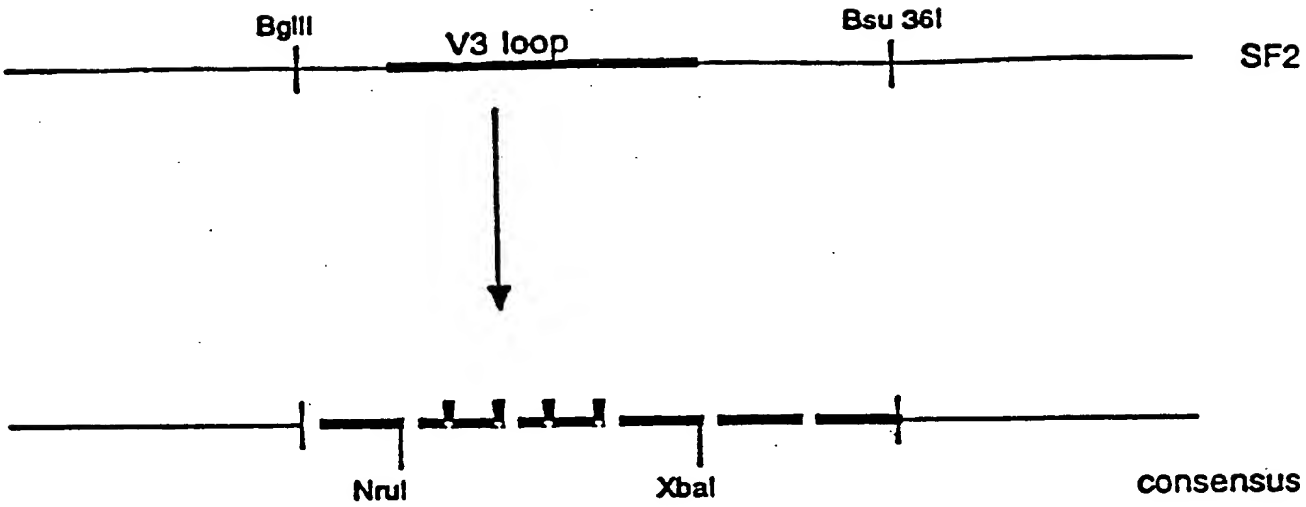

FIGURE 4

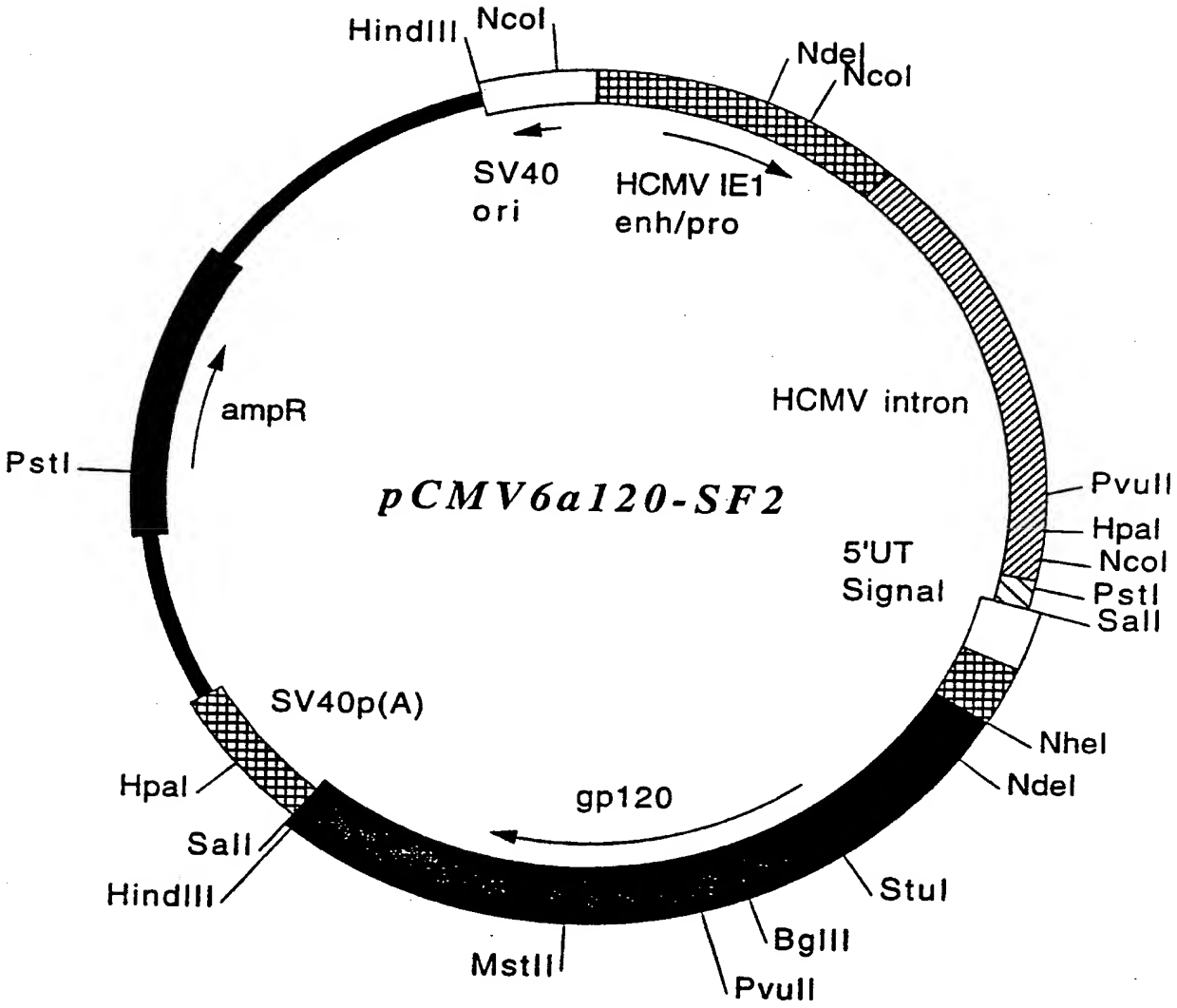

FIGURE 5

```
Consensus                .T.VTGG.AARTT.G..SLF..G.SQ.IQLI      31

re3  (15-45)             A.YA..AAQGHA.NSFV...RS.A..NLK.V      45
hcj7.pep (15-45)         S.Q....Q..H.VR.VA.I.SP.SR.D.S..      45
ny1.pep (15-45)          S.R....QQG.AVH.IA...SL.A..K...V      45
GE11.2  (15-45)          S.H.M.AQQG.VAK.FT...GP.PA.K....      45
s71957.p (15-45)         S.H...AVQGHSIR.LT...TS.PA.K...V      45
ec10  (15-45)            E.H....I..K..ASLTG..NL.AK.N....      45
sp2.tc  (15-45)          E.H....N.G.AAA.IAG..TL.AK.NV...      45
M2.2 (15-45)             Q.R....T..QS.ARIAG..SL.AR.N....      45
rela  (15-45)            Q.H.M..T.G.NAY.LT.FLSV.A..K....      45
168  (15-45)             E.H.M..A.SS..YRFA...TS.PA.K...V      45
M1.5 (15-45)             E.H....S..S..ATFSK..MP.A..N....      45
hpcprc1a.p1 (14-44)      G.TRV..A.....SSFA..LTH.P..N...V      44
gm2.tc  (15-45)          G.H....A...DAFRFS...TR.P..N....      45
s71864.p1 (15-45)        A.NM...AP....YKLTT..SY.A..K....      45
i15.tc  (15-45)          HNH....TS..N.F.ITT..TQ.P..KL..V      45
sp1.tc  (15-45)          G.H....A...NAHSLT..LAP.A..K....      45
re37b  (15-45)           T.R.S..T..H..A.LT...SP.PR.N.H.V      45
re39  (15-45)            T.H.S..T.G...ASLT.F.AP.A..R...V      45
hpchcj2.p (15-45)        T.H....ATGH..S.IA...LP.A..K....      45
hpccgenom.p1 (15-45)     D.YAS..AQG.S.L.FT...TP.A..K....      45
hpcprc4a.p2 (14-44)      D.YAS..A.G.A.Y.IT...AP.A..N....      44
hpcprc11a.p1 (4-34)      R.YAS..A.G...H.FT...ST.AR.N....      34
re72b  (15-45)           Q.Y....K..Q.VS.FTG..SS.P..K....      45
hpcprc3a.p1 (4-34)       D.Y.S..A...SIS.FT...TP.A..K...V      34
re70  (15-45)            S.Y....E.S...R.FA...TL.S..K....      45
re42  (15-45)            N.Y....S.G.AVA.FAG.LQP.AK.NV...      45
63  (15-45)              H.R....QV.FR.H.LV...TQ.P..K...V      45
hcvj  (15-45)            H.H....RV.SS.QSLV.WLSQ.P..K...V      45
hcvj1  (15-45)           H.R....VQGHV.STLT...RP.A..K...V      45
re38  (15-45)            N.R....VQG.D.S.LV...SL.P..K...V      45
bk  (15-45)              D.H....AQ.K..NRLV.M.AS.P..K....      45
re5  (15-45)             E.H....AS....QRFT.F.DL.P..K...V      45
re35  (15-45)            T.YM...AN....Q.FV...TP.PA.K...V      45
re36  (15-45)            E.H....TS....Q.FV...SA.A..K...V      45
re54  (15-45)            G.H....Q.....QSFT...SP.PQ.K....      45
re56  (15-45)            R.H....K..H..K.FA...TP.P..N....      45
64  (15-45)              E.R...AVQGHGAL.LA...TP.P..K....      45
re62  (15-45)            E.R...AI.G..ASSFAG..TS.A..K...V      45
re41  (15-45)            E.R...QQVG...QSLT...TP.P..T....      45
hpcvjk3.p (15-45)        Q.R...AQVG...SSLT...TP.P..N...V      45
jk2  (15-45)             R.Q...AQ.GH..S.LA...TP.P..K...V      45
hcv1  (15-45)            E.H....S.GH.VS.FV..LAP.AK.NV...      45
us5.tc  (15-45)          E.H....S.GH.VT.IA...TS.AK.N....      45
i21.tc  (15-45)          S.H....T..H.VA.FS...TV.PK.N....      45
M3.1 (15-45)             E.H....A..Y.AA.LA...TS.AK.N....      45
H77  (15-45)             E.H....S.G...A.LVG.LTP.AK.N....      45
re43  (15-45)            G.H....S.G.A.A.IAG.LTP.AR.N...V      45
Ge6.3  (15-45)           K.H....S.....S.IA..LTP.AK.NV...      45
CO11  (15-45)            K.Y....SQ.QA.F.FT..LQQ.AK.N....      45
TH  (15-45)              E.T....S..HGAL.IA...NQ.AR.N....      45
O11S  (15-45)            E.Y....AS..S.FTLVG..KQ.SQ.N...V      45
re4a  (15-45)            Q.Y.S..SSG...S.LV.I.SP.A..NL...      45
q1  (15-45)              E.Y.S..A..Q..ARFAGF.QS.AK.N....      45
q3  (15-45)              E.Y.S..S..Q..A.FVR..ET.PK.N....      45
gh1.tc  (15-45)          S.Y.S..AQ..AAQ.IT...SR.S..K...V      45
M4.1 (15-45)             S.Y....TQG.AAS.LT...SA.A..N....      45
```

FIGURE 6A

nac5.tc (15-45)	N.Y.S..T.GH.GH.LTA..SP.A..N....	45
hpcgenanti.p3 (15-45)	S.I.S..TV....HSLA...TQ.A..K....	45
hcj4 (15-45)	E.YTS..A.SH..STLA...SP.A..R...V	45
hpchcv.p2 (15-45)	H.LT...H...L.S.FAG..TP.P..R....	45
hcj1 (15-45)	E.I.S..Q...AMS.LV...TP.AR.N....	45
hct18 (15-45)	E.YTS..N.GH.MT.IVRF.AP.PK.NVH..	45
hct27 (15-45)	T.YT...N.....QALT.F.SP.AK.D....	45
hcve1 (15-45)	E.YT...ST....Q.LV...SR.AK.D....	45
ge12 (15-45)	A.YTS..S.....Q.FA...SL.SQ.K...V	45
LG (15-45)	A.YT...SV....H.FS...SQ.AK.N....	45
jt.p3.x (15-45)	V.YT...SQ..H.QSVT.F.TQ.PA.R....	45
us4.tc (15-45)	H.YT...TV..S.Q.LVGFLSP.P..N....	45
jk1 (15-45)	T.Y.SV.H.SQ..RRVA.F.SP.SA.K...V	45
hpcvjk4.p (15-45)	T.T.S..H.SQI.R.VT.F.SP.SA.K...V	45
hpce2cor.p (15-45)	K.SL..VTR..AAARLTA..SS.P..R....	45
hpcns34d.p (15-45)	G.SL...AR..AAS.LAG..SS.P..R....	45
FTO.1 (15-45)	V.Q.SPPQ.GY..SVLTGILSP.AK.N...V	45
Gj6.1 (15-45)	V.Q.S..Q.GY..SVLTGILSP.AK.N...V	45
re7 (15-45)	G.YTV..AS.F..SRLT...AL.P..R...V	45
hcvkf (16-46)	N.HTV..TEGFA.QRLT...AL.P..K....	46
arg2.tc (15-45)	S.RTA..AQ.FN.Y.VA.I.SP.P..R...V	45
hcj6 (15-45)	Q.HTV..ST.HNARTLTGM.SL.AR.K....	45
hpchcj5.p (15-45)	N.RTVA.S..A..R.FT.M.SS.SK.NL...	45
rs1.pep (15-45)	Q.RTV..QVGHSVR.FT...SA.SA.N....	45
re71 (15-45)	E.HT..AVSGH..NVLT...SS.S..N....	45
re6 (15-45)	V.RT..EV....ANTFA...TT.P..N..I.	45
hcj8.pep (15-45)	T.YSS.QE.G..VA.FAG..TT.AK.NLY..	45
re40 (15-45)	S.S.V..RQ.SA.FRFT.F.SR.PT.E.K..	45
hpcencr.p (15-45)	N.YT.A.SM.QSIYRLTDI.ST.P..KL..V	45
re55b (15-45)	R.ILMA.RQ.EV.QSFPG..SLAP..K.H..	45
aus1.tc (15-45)	D.YA...SV.SIMA.IARF.SP.AR.D....	45
PC2.1 (15-44)	E.YA..AS.GHDVSSFAR..AP.AR.N...-	44
hct23 (15-45)	E.HR...S...S.A.VA...TP.AR.N....	45
re34 (15-45)	N.RAV.MVQS...YALT...DS.AA.KL..V	45

FIGURE 6B SUBSTITUTE SHEET (RULE 26)

INTERNATIONAL SEARCH REPORT

International Application No
US 94/05166

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/40 C12N15/48 C07K14/16

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JOURNAL OF VIROLOGY vol. 66, no. 1 , January 1993 pages 399 - 404 S. LI ET AL. 'Influenza A virus transfectants with chimeric hemagglutinins containing epitopes from different subtypes' *see the whole article and more particularly, page 400, right column and page 402, right column, last paragr4ph*	1,2,4, 17,22,33
X	EP,A,0 448 095 (H. J. WOLF ET AL.) 25 September 1991 *see the whole document and more particularly, page 3, lines 10-14, page 4, lines, 6-8 and 45*	1-37

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
5 October 1994	27 -10- 1994

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax (+ 31-70) 340-3016	Marie, A

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP,A,0 449 116 (H. WOLF ET AL.) 2 October 1991 *see the whole document*	1-37
X	JOURNAL OF GENERAL VIROLOGY vol. 74 , 1993 pages 1261 - 1269 R. WAGNER ET AL. 'Induction of cytolytic T lymphocytes directed towards the V3 loop of the human immunodeficiency virus type 1 external glycoprotein gp 120 by p55gag/BV3 chimeric vaccinia viruses' *see the whole article*	1-37
Y	AIDS vol. 7 , 1993 pages 759 - 767 J. BLOMBERG ET AL. 'A survey of synthetic HIV-1 peptides...' *see the whole document*	1-37
Y	FR,A,2 677 363 (INSTITUT PASTEUR ET AL.) 11 December 1992 *see the whole document*	1-37

1

INTERNATIONAL SEARCH REPORT

...formation on patent family members

Inter... nal Application No
P... US 94/05166

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP-A-0448095	25-09-91	NONE	
EP-A-0449116	02-10-91	NONE	
FR-A-2677363	11-12-92	NONE	